The evolution of Health & Place: Text mining papers published between 1995 and 2018

The evolution of Health & Place: Text mining papers published between 1995 and 2018

Mark Green (University of Liverpool)
Michael Widener (University of Toronto)
Frances Darlington-Pollock (University of Liverpool)
Jamie Pearce (University of Edinburgh)

Abstract
Over the past 25 years, research published in Health & Place has led the way in revealing how and why place-based processes are implicated in a broad range of health-related concerns. We examine trends in terms published in the journal between 1995 and 2018. Terms used in the journal have diversified over this period, including a decline in 'health' overall, as well as greater usage of 'diet', 'alcohol' and 'depression'. There was a noticeable increase in using 'neighbourhood' to describe context. Topic modelling reveals that papers are increasingly concerned with specific aspects of place (e.g. the built environment). Overall, the findings emphasise that Health & Place has matured into the leading outlet for examining the geographical dimensions of health with important and enduring insights that continue to inform research and policy.

Keywords
Text; text mining; health; health geography.

Word count
5652

1.0 Introduction

To mark the 25th anniversary of Health & Place, we examine trends in the types of work published in Health & Place between 1995 and 2018. Our paper complements the accompanying bibliometric analysis also celebrating the milestone (Moon and Pearce, 2020). Through undertaking a text mining approach to assess how terms and themes within articles have changed over time, rather than the citations or author origins of papers as Moon and Pearce focus on, we provide a fine grained analysis of the evolution of the content published in Health & Place.

Over the past quarter of a century, articles appearing in the pages of Health & Place have been at the forefront of revealing the ways in which places matter for health (Moon, 2008). Health & Place originally emerged at a time of flux across the social sciences and sought to consolidate synergies with researchers interested in health-place connections in cognate disciplines including Geography, Sociology, Public Health, Social Policy and others (Moon, 1995). The interdisciplinary outlook of the journal, aligned to sub-disciplinary developments (e.g. Kearns, 1993; Moon, 1995; Kearns and Moon, 2002), and papers appearing in Health & Place have developed rapidly in three interrelated ways. First, authors of the journal's papers have adopted a broader range of theoretical standpoints, which has resulted in greater conceptual plurality in the published research. These changes reflect in part the evolution in the social sciences from a medicalised focus on disease and healthcare to a theory-led and enriched analyses of the interactions between health and place (Kearns and Moon, 2002). This theoretical expansion has helped to facilitate the adoption and development of a greater diversity of methods used in answering key questions about the role of place on health. Second, the past 25 years have also seen a wider set of disciplines pursue an interest in the connections between heath and place. Scholars in fields including Epidemiology have provided new insights into these questions adding to the theoretical richness, deepening the methodological toolbox and enhancing analytical rigour. Finally, and as this paper examines comprehensively, there has been a broadening of the thematic focus of the published research from a traditional focus on environmental determinants of disease and the role of health service provision, to a wider set of concerns covering many of the largest public health challenges including non-communicable disease prevention, mental health and health inequities (Richardson *et al.*, 2013). Further, researchers are increasingly grappling with a wider range of places including neighbourhoods and broader activity spaces, primary and secondary healthcare settings, schools, prisons and other contexts important in shaping people's health-related experiences.

To do this, we turn to methods born out of the big data revolution. While much of the interest in big data have focused on the size of datasets, new forms of data types offer exciting opportunities for studying the (geographical) determinants of health. Traditionally, quantitative data (e.g. surveys, administrative databases, Censuses) are often numerical and structured in design to help facilitate statistical analyses (Connelly *et al.*, 2016; Timmins *et al.*, 2018). The growing capabilities of computers, alongside newly developed algorithms, have repurposed unstructured data formats allowing their integration into quantitative analyses.

Historically, humans have stored information in the form of text through books, journals and written records. There is increasing realisation that these vast quantities of textual records might be mined to process the information and generate new insights. Examples of text mining within health-related research include: automatic coding of discharge summaries

(Kukafka *et al.*, 2006); mining electronic health records to identify the clustering of co-morbidities and how they relate to treatment procedures (Metsker *et al.*, 2017); classifying disease phenotyping for diseases from medical records (Ning *et al.*, 2019); developing predictive algorithms for cancer based on pathology and radiology reports (Spasić *et al.*, 2014). Many of these applications are in biomedical fields and are rarely used in by researchers interested in the connections between health and place. We are only aware of Porter, Atkinson and Gregory (2015) who examine spatial patterns in themes reported in Registrar-General reports for Great Britain between 1850 and 1911.

Text mining approaches have also been used to process research published in journals to understand how fields have evolved. Examples include: identifying how the usage of terms changed in old medical periodicals (Ferry, 2015); exploring the evolution of research relating to topics (Balan, Gerits and Vanduffel, 2014); finding terms relating to subfields of research (Westergaard *et al.*, 2018). Such approaches can supplement traditional literature reviews through generating data driven insights that are less intensive, faster, reproducible, output quantifiable information, and can scale with the increasing quantities of complex text data.

2.0 Methodology

2.1 Data

All papers related to research published in Health & Place between 1995 and 2018 (volumes 1 to 54) were electronically downloaded from the journal's website. We include all article types comprising research articles, short reports, opinion pieces and commentaries, editorials, book reviews and review articles. We excluded editorial board pages, referee lists, volume contents, keyword and author indexes. Articles were downloaded as PDFs since this was the only consistent file type available across all years.

PDF files for each year were joined into a single year file and then converted to text format using Adobe Reader Pro. Conversion from PDF to text format can be messy (Westergaard *et al.*, 2018) and required manually editing files to remove non-alphanumeric characters that were not present in papers (using search and find functions in Adobe Reader Pro). All headers were removed to avoid repeating 'Health & Place' throughout. For the topic modelling analyses, we read the PDF directly into R statistical software and converted individual files to text format.

Text files were converted to a corpus which is a database for text information that allowed the text data to be processed into numerical information (Silge and Robinson, 2017). Data cleaning of the corpus included removing all special characters (i.e. anything that is not an alphanumerical symbol or space), converted all terms to lower case (otherwise terms with capital letters would be considered as different to their lower case spelling), removed 'stop' words (e.g. the, of, in etc which add little value to understanding the language of research), removed all numbers, removed all whitespace left and finally converted all words to their root (e.g. remove 'ing', 's', 'ed' etc). The corpus allowed the creation of a document-term matrix identifying how common terms were across each year.

2.2 Statistical Analysis

Natural language processing and machine learning offer the potential for processing vast quantities of electronic text information (Silge and Robinson, 2017). We use a combination of descriptive and formal techniques to evaluate trends in terms throughout the journal's publication history. Term frequency is explored over time using summary statistics and we focus on several case studies that highlight key features of debate during this period. (selected

by author's expert opinion). We consider terms independently, as well as common combinations of terms as well. Sentiment analysis is undertaken through classifying non-neutral words as 'positive' or 'negative' (using the bing lexicon) to understand the emotion of terms used throughout the journal (Liu, 2012). Finally, a Latent Dirichlet Allocation model was used to classify papers through identifying a latent structure of unstructured terms within papers (Blei, Ng and Jordan, 2003).

3.0 Results

Table A1 (Appendix) presents summary statistics for the papers selected within our sample. Between 1995 and 2018, the journal published a total of 19 312 870 words from 804 372 sentences and 1928 articles. The general trend demonstrates a clear increase in the number of words published from 1995 to 2012, before declining thereafter. At its height in 2012, the journal published ~9 times as many words (1 691 656) and ~8 times as many papers (171) annually than compared to when it was first established in 1995 (192 913 and 22 respectively).

3.1 Health outcomes

We first consider trends in the most common term 'health'. Figure 1A presents the term frequency for 'health' which is the proportion of occurrences of a term across the entire set of documents. The use of the term 'health' displayed a small increase in frequency initially between 1995 and 2003, however has become less popular over time especially since 2012. While this may appear counterintuitive for the journal, it can be explained by the evolution of the field. Earlier papers published focused on identifying *whether* geographical context mattered for understanding health. Many papers examined this through focusing on general outcomes such as self-rated health status. With the association between health and place established, research changed focus towards understanding *how* geography mattered, in addition to examining a greater diversity of topics and more specific outcomes (Richardson *et al.*, 2013). For instance, terms such as 'wellbeing' have become more popular in occurrence since 2010 (term frequency increased from 0.0004 to 0.0013 between 2010 and 2018). This is arguably a natural progression once the framework of place as a means to understand health been established.

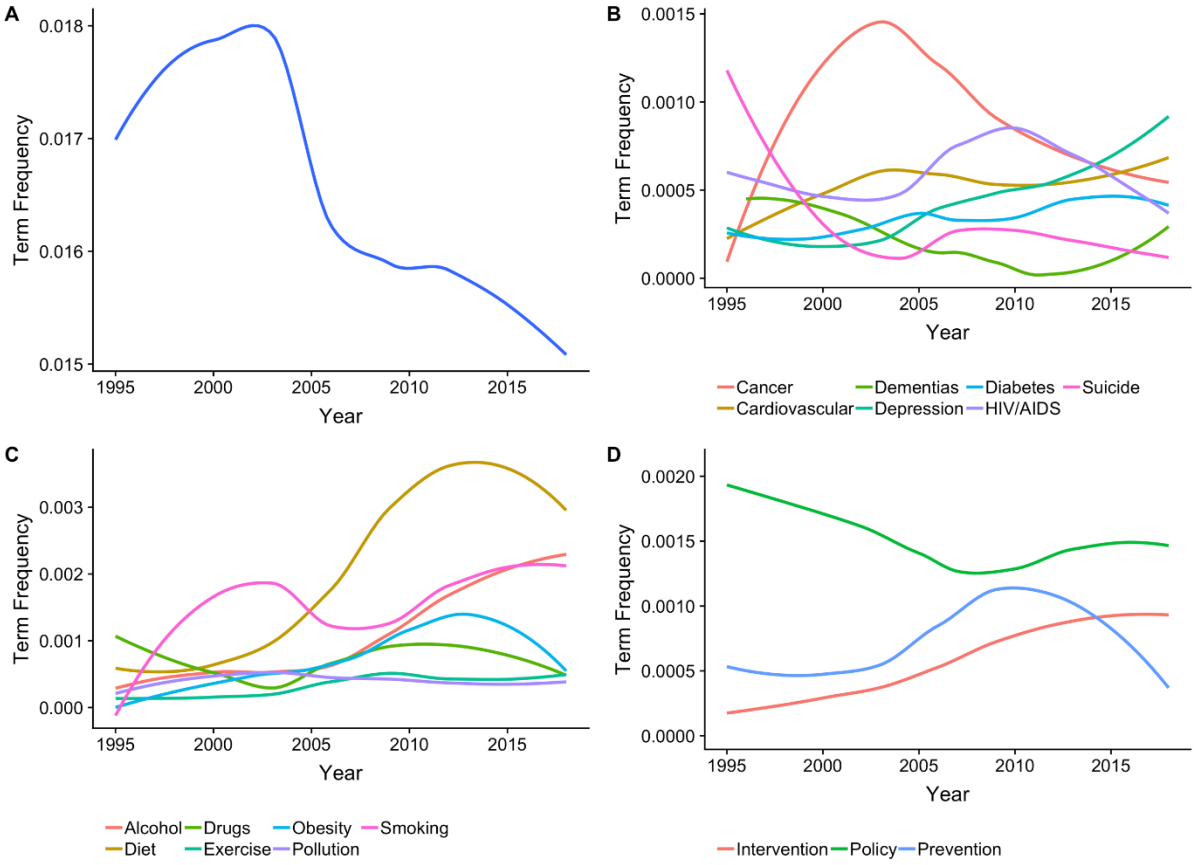

Figure 1: Smoothed term frequency trends in Health & Place, 1995-2018 for A: Health, B: Health Outcomes, C: Risk Factors, D: Policy.

We next examined trends in specific health outcomes. We selected seven outcomes that represented the largest contributions to several metrics from the Global Burden of Disease study including mortality rates, years of life lost and disability adjusted life years globally (GBD, 2017). Trends in occurrences of these conditions is presented in Figure 1B.

The ordering of term frequency does not necessarily match the prevalence of these diseases reflecting the diversity of interests published in the journal. Suicide was the most common of these terms at the inception of the journal, but has declined sharply since and is the least popular of the terms by the end of the period. Cancer quickly becomes the most common issue and remains as such for most of the period peaking in 2003, before declining sharply thereafter. HIV/AIDS sees an increasing popularity in usage during the 2000s before declining rapidly from 2010. By the end of the period, depression is the most common outcomes (whereas it had started the period as a relatively uncommon term), reflecting the growing interests in mental health research over the period. Finally, dementia-related conditions have shown growing interest since 2011 suggesting a future important issue in the journal. This interest is likely to continue with the growing prevalence of such diseases in ageing populations.

Trends in terms relating to risk factors and exposures were examined (Figure 1C). We selected the seven risk factors that had the largest influence on ill health and disability (GBD, 2018). Diet (and related terms i.e. food and nutrition) was the most common term for the majority of the period having overtaken smoking-related terms in 2005. The usage of diet-related terms is far higher than the other terms we considered, although since 2013 they have

declined in occurrence (a similar trend is observed for obesity which is likely related). Smoking was most important prior to 2005 and remains a common term with a curvilinear trend post 2005. Interestingly, e-cigarettes (not covered here) is first mentioned in 2014 however by 2018 is just as common as obesity, which suggests an emerging area for the journal. Alcohol also becomes an increasingly more prevalent term over the period as well.

Finally, we focused on terms relating to policy and applications of research (Figure 1D). Both 'policy' and 'prevention' have declined in usage over the period, contrasting to an upward trend in use of 'intervention'. 'Intervention' was not a common term at the start of the journal and is likely a phrase that has gained in popularity due to the growing diversity of fields (particularly epidemiology and medical sciences) publishing in the journal that have introduced the term to the readership, as well as changes in the funding landscape which have encouraged evaluative work. The evolution towards intervention-led research may also reflect maturing of the field through identifying specific geographical phenomenon that can be altered to improve health. This is likely a natural progression of research following the establishment of how and why geographical context matters.

3.2 Describing geographical context

Figure 2 focuses on trends in varying terms that correspond to how geographical context is described in papers in the journal. It reveals nuanced insights in how the language of Health & Place has evolved over time. 'Neighbourhood' was a fairly rare term used in the early years of the journal, however it has increased in popularity over time and is now the most common of the terms considered. The increasing popularity coincides with the development of the subfield of neighbourhood effects. Terms that have become less popular include 'area', 'community' and 'geography' (all popular at the beginning of the journal). It may be that 'area' and 'geography' are more general phrases than compared to 'neighbourhood' hence their respective trends. This explanation does not apply to 'community', which may reflect a shift of interests away from studying the topic or a substitution of terms. The decline in 'geography' may also represent the growing diversity of fields publishing in Health & Place.

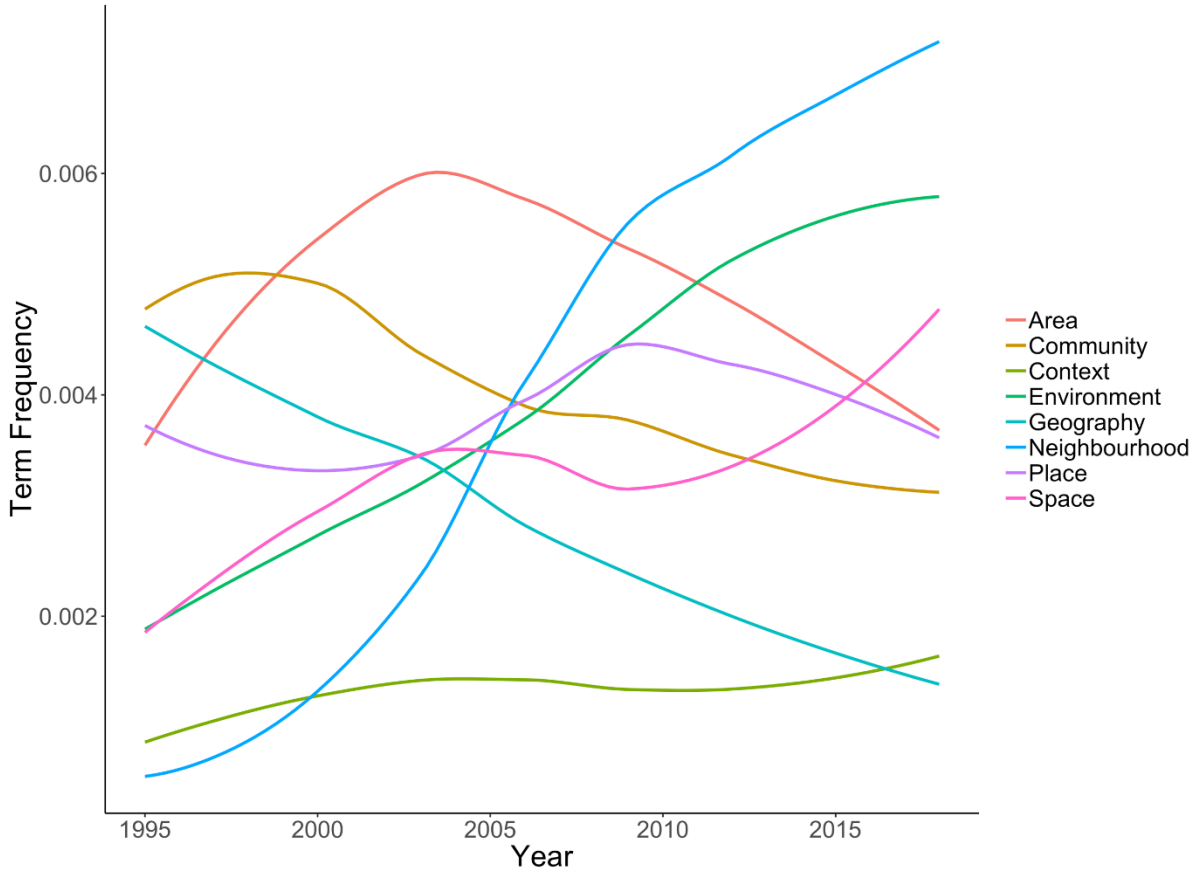

Figure 2: Smoothed term frequency for words relating to geographical context in Health & Place, 1995-2018 (note: terms including their variant spellings).

3.4 Methods

We next compare trends in specific methodological approaches. There is no systematic list to define what are the most popular amongst studies concerned with geographical aspect of health. We selected eight methods that give a good range of qualitative and quantitative approaches, as well as approaches that we anticipate will be increasingly adopted amongst researchers in the field (Figure 3).

Interviews were the most common of the terms we considered and have seen rising prevalence in occurrence over the period. The only other method which also saw a large increase in usage over the period was regression. Both GIS and multi-level models saw growth and decline over the period, albeit peaking in 2006 and 2010 respectively. The other methods were less common over the period, but all saw rising trends since 2012 and may represent some future directions in the field. This included agent based models which gained popularity at the beginning of the period as well.

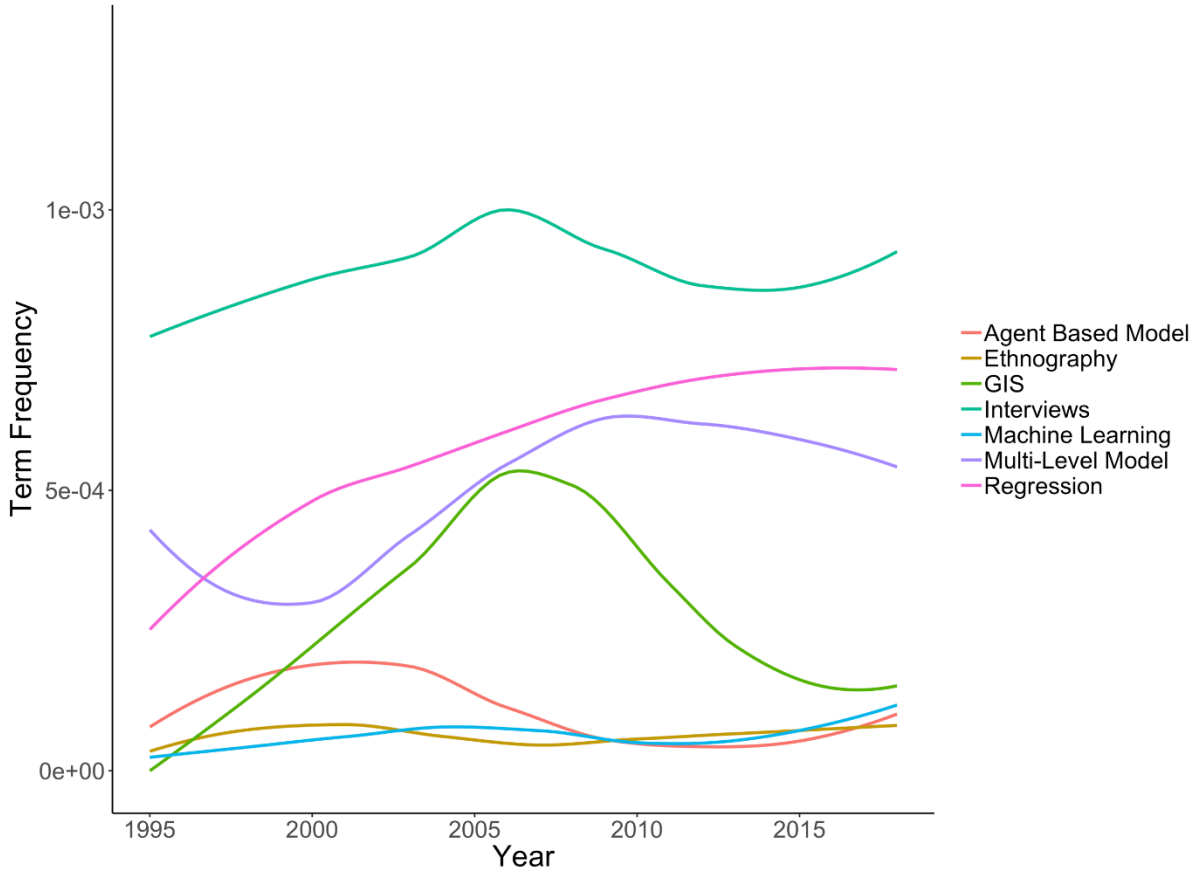

Figure 3: Smoothed term frequency for methods in Health & Place, 1995-2018 (note: terms including their variant spellings).

3.5 Topic modelling

Using the 1928 papers from 1995-2018, we sought to classify papers based on the terms within each to generate a classification of topics published in the journal. A Latent Dirichlet Allocation model was used to classify papers based on the presence of terms used within them (Blei, Ng and Jordan, 2003). The number of clusters was identified through comparing model fit statistics, as well as a subjective judgement based on the interpretation of the results. We aimed to identify the parsimonious solution which preserves the diversity in article types, while also seeking to minimise the complexity of having a large number of clusters. A nine cluster solution was selected as the most appropriate. Figure A2 (Appendix) plots the probabilities (beta values) for the top 15 most common terms in each cluster. Descriptions of each clusters (including their estimated prevalence) are provided in Table 1. We have contextualised their interpretation through including the paper with the highest associated gamma value reflecting the paper which is the best match for the cluster (all values were at least 99.99% match).

Cluster	Description	Example	Prevalence
1	The topics reflected research focusing on the role of social deprivation and how it related to health outcomes. Research included investigations of social	Tunstall, Cabieses and Shaw, 2012: A study of how migration patterns influence health inequalities by neighbourhood deprivation.	11.1%

	determinants of health, as well as ecological studies.		
2	The cluster was only one of two clusters where 'health' was not the most popular term. 'Care' was the most popular term, with other common terms reflecting health services research ('hospital', 'services', 'patients', 'medical', 'access'). Studies often explored barriers to care services utilisation, including accessibility.	Kümpers *et al.*, 2006: The study compared dementia policies and approaches for integrated care between four areas in England and the Netherlands.	10.7%
3	The cluster was similar to the fifth cluster through focusing on terms relating to the built environment ('environment', 'built', 'physical', 'activity'), however differed by focusing more on studying the built environment rather than necessarily it's association to specific health outcomes.	Cusack *et al.*, 2017: A study on how neighbourhood green space was associated to birth weight was the strongest match to the cluster.	11.2%
4	The most popular terms for the cluster reflected spatial analyses of health outcomes (particularly 'cancer', 'mortality' and 'HIV'). Many studies were ecological and involved clustering based methods.	Jongsthapongpanth and Bagchi-Sen, 2010: An analysis of the spatial clustering of AIDS in Thailand.	9.0%
5	'Health' was not the most popular term, but was superseded by 'food', 'activity' and 'physical'. Terms reflected built environment research, often focusing on understanding their associations to obesity or physical activity.	Audrey and Batista-Ferrer, 2015: A systematic review of interventions designed focused on improving the built environment to improve health in children and adolescents (particularly obesity, active travel and road safety).	14.3%
6	The beta values for the topic were the largest suggesting the cluster was well defined in the model. 'Social' is very high and almost as strong as 'health'. The other terms suggested social capital ('social' and 'capital' were both important, which was reflected from exploring the papers included in the cluster), as well as descriptors of geographical context	Elgar *et al.*, 2011: An exploration of the association between social capital to health and life satisfaction across 50 countries.	11.4%

	('neighbourhood', 'place', 'community') which are linked to the application of social capital.		
7	The most popular terms that defined the cluster were similar to cluster 6, but were subtly different in focus. The interest in studying geographical context was key ('place', 'community', 'home'). However, papers were more focused on exploring qualitative understandings of place (and how we engage with them).	Coleman and Kearns, 2015: The study used photo-elicitation interviews to explore the role of bluespace in ageing and wellbeing.	11.8%
8	Beta values were fairly low for the cluster suggesting that it was less distinct overall. The most common terms were somewhat general, though included phrases conceptualising context ('place', 'public', 'local'), as well as 'policy'. Papers were mixed and likely reflected those that did not fit the other clusters well (given the diversity of the journal).	Dooris, 2013: They used interviews with policy makers and experts to examine how health promotion messages are designed.	10.4%
9	The topics covered in the papers classified in the final cluster reflected approaches to quantitatively model ('data', 'model') social factors ('social', 'education', 'school') on child health and health-related behaviours ('smoking', 'tobacco').	Fotso and Kuate-Defo, 2005: The study used multi-level modelling to identify the role of community and family socioeconomic status on child health.	9.9%

Table 1: Summary of clusters identified from classifying papers published in Health & Place, 1995-2018

We also repeated the analysis by comparing the first (1995-1999; n = 125) and last (2014-2018; n = 665) five years of the journal to identify how the types of papers have changed. For both models, a six cluster solution was selected as most appropriate. Descriptions of the clusters for the period 1995-1999 are given in Table 2 (also see Figure A3 in the Appendix). Table 3 (also see Figure A4 in the Appendix) presents the results for 2014-2018.

Comparing the results between 1995-1999 and 2014-2018, the topics of papers have evolved reflecting the maturing of the field. Between 1995 and 1999, the field was focusing on understanding if geographic context was important for understanding health. When these associations became more established and consistent, papers began to move towards exploring specific aspects and mechanisms of place (and concepts of health). This was helped both by the research set up earlier in the journals formation, as well as the availability of

newer more detailed data sets. This can be seen in the cluster characteristics that are far more specific in the later period. The term 'care' was also common in all clusters between 1995 and 1999, whereas was less important in 2014-2018. It is also noticeable how the papers provided as examples of the clusters are all bar one single author in 1995-1999, a contrast to more recent papers which reflect greater collaboration and research teams.

Cluster	Description	Example	Prevalence
1	Studies focusing on HIV/AIDS ('hivaids', 'aids') were common in the cluster and reflected the focus of many papers. Other common terms included health services ('pharmacy'), context and interconnections between people ('community', 'network'). These were typically explored in relation to HIV/AIDS to understand experiences of the disease and opportunities to accommodate them.	Wilton, 1996: The qualitative study explored experiences of men living with HIV/AIDS on everyday life in Los Angeles.	12.8%
2	The terms in the cluster were mixed and general, suggesting it contains papers that were not classified elsewhere. In the formative years of the journal this may not be surprising as it will have attracted a diverse range of topics while subfields begin to develop.	Congdon, 1997: A Bayesian spatial analysis exploring the social drivers of small area suicide rates in London.	16.0%
3	The terms 'community' and 'place' reflected that papers in the cluster examined how conceptualisations of place and space with respect to health (with other terms being their focus of these discussions). They were important early papers in establishing what place-based research was.	Madge, 1998: An analysis of therapeutic landscapes in The Gambia to understand the role of place and cultural context in shaping health care practices	16.0%
4	The term 'mental' was common in papers in the cluster. The cluster is largely picking up the special issue 'Space, Place and the Asylum' in 1997 (as well as related papers preceding and following it).	Philo, 1997:Philo argues for research into 'asylum geographies' through reviewing the literature on the role of mental health facilities	16.8%

5	Papers utilising small area health statistics to explore their spatial patterns. These included analyses of the demographic and socioeconomic correlates to health patterns.	Congdon, 1999: The study analyses small area data on socioeconomic status to understand health care utilisation patterns in London.	22.4%
6	The terms in the final cluster represent health services research (including the topics they have investigated). They include access, utilisation and the experiences of services.	Lewis and Rapaport, 1995: An examination of health services pressures in Pacific Island contexts particularly with respect to alcohol, HIV/AIDS and environmental issues.	16.0%

Table 2: Summary of clusters identified from classifying papers published in Health & Place, 1995-1999

Cluster	Description	Example	Prevalence
1	The cluster is defined by popular terms including 'mental', context ('neighbourhood', 'place') and quantitative study terms ('data', 'model'). 'Social' and 'capital' are also frequent and investigations into the role of social capital is a common topic among papers.	Cheung, 2014: The study examined the role of social ties in promoting wellbeing adolescent migrants in China	15.0%
2	The cluster is themed around the built environment mainly focusing on associations to alcohol and smoking (separating it out from clusters five and six).	Decker *et al.*, 2018: The study is a systematic review of neighbourhood characteristics that were associated with adolescent health and health-related behaviours.	17.3%
3	The cluster was mostly concerned with health services and community care research. There was a wide breadth of topics investigated with these themes, particularly how care is experienced as well as links to space and place	Coleman and Kearns, 2015 (see Table 1)	25.1%
4	Accessibility to health services and small area statistics were the most common themes present in the papers classified in the cluster. Many studies were ecological analyses.	El Anshasy and Katsaiti, 2015: The paper used panel data to model how natural resource expenditure was associated to health care spending and health outcomes.	14.4%
5	The terms which were popular in the cluster referred to	Audrey and Batista-Ferrer, 2015 (see Table 1)	13.1%

	walkability, built environment, physical activity and children		
6	The only cluster where the term 'health' is not the most popular term. Instead, 'food' is more common and this is reflected in the other terms which relate to foodscapes and obesogenic environments	Lytle and Sokol, 2017: It was a systematic review of how the food environment was measured.	15.0%

Table 3: Summary of clusters identified from classifying papers published in Health & Place, 2014-2018

4.0 Discussion

In this paper, we've considered the evolution of research within Health & Place between 1995 and 2018. We find growing numbers of papers and words published over time with the expansion of research covered in the journal. The growth of research reflects a greater diversity of topics published including a decline in the use of the term 'health'. Health outcomes of focus have changed, with less interest in cancer and growing research focusing on diet and mental health. There have been subtle changes in the language used throughout, particularly a decline in the use of 'geography' in favour of 'neighbourhood'. We define 10 'types' of papers that characterise the research published in the journal and stratifying the analyses over time reveals articles types are becoming more focused on specific aspects of neighbourhood effects.

The growing diversity of research published in Health & Place demonstrates the maturing of the interdisciplinary fields (and sub-disciplines) represented in general. Initial discussions and research were tentative in the formative years of the journal, representing the early development of the field. Following early calls prioritising 'place' in health-related research (Kearns, 1993), many papers began to define what exactly place-based research might look like (Kearns and Moon, 2002), before exploring whether geographical context mattered overall. With growing confidence that place mattered, studies have evolved to drill down into specific features of geographical context that may influence health. This was evident in the changes over time of the topic modelling analysis with studies examining the role of the built and food environment becoming prominent in the most recent years compared to the formative years of the journal.

Interesting insights are revealed about subtle changes to language regards geographical context. The term 'neighbourhood' has come to the forefront in the journal despite being rarely used at the journals conception. The growth in the term seems to have appeared at the expense of terms such as 'geography', 'area', 'community' and 'place' which have all declined over time. These trends have coincided with increasing interest in identifying neighbourhood effects on health (Oakes *et al.*, 2015). The use of terminology might be explained through moving away from vague terms towards more specific mechanisms. The decline in the term 'community' is an exception to this understanding. A wide variety of research has investigated the importance of communities for health (Arcaya *et al.*, 2018). Our work suggests the need for greater investigation therefore into the reasons behind this decline

and potential a renewed focus onto community systems amongst studies on the relations between health and place.

Our study is one of the few applications of text mining and natural language processing in geographical work on health concerns and serves to demonstrate the potential of such approaches. The growing diversity of new forms data combined with advances in analytical algorithms (Timmins *et al.*, 2018), opens up new opportunities to understand the geographical determinants of health. Text mining may also help to supplement qualitative approaches towards reviewing the literature through providing quantifiable data, as well as the ability to process larger quantities of information.

There are several limitations to our study. We only consider papers published in Health & Place and therefore we cannot claim wider generalisability to the field of Health Geography given that academics publish in a broader range of journals. While Health & Place has a defining role in the evolution of this research area, it is only one lens through which we view broader trends in the field. Future research should look to strengthen this investigation to improve our understanding of the field. Text data is messy and the conversion of PDF to text files was limited when trying to convert tables and figures from papers that were difficult to clean (Westergaard *et al.*, 2018). We were unable to extract information from all tables and examine how trends have varied in the results presented in papers. Analysis of papers does not consider the quality of papers or relate trends to outcomes (e.g. citations) that might help to contextualise and prioritise the papers. Finally, our study only considers the words as independent terms and ignores the wider context of language that they were situated within (e.g. terms being used sarcastically).

In conclusion, our study demonstrates the evolution of Health & Place into a platform for diverse and internationally leading research into the geographical determinants of health. We fully expect the field to remain dynamic and continue to reinvent investigations between health and place. The 25 years of research published in the journal marks an exciting maturing of the field and we look forward to discovering where the next 25 years might lead.

Acknowledgements
Mark Green was funded by an 'Early Career Researcher Award' by the University of Liverpool to complete the research. We would like to thank the anonymous reviewer for their helpful and positive comments which have improved the quality of the paper.

References

El Anshasy, A. A. and Katsaiti, M.-S. (2015) 'Are natural resources bad for health?', *Health & Place*, 32, pp. 29–42. doi: https://doi.org/10.1016/j.healthplace.2014.12.011.

Arcaya, M. C., Schnake-Mahl, A., Binet, A., Simpson, S., Church, M. S., Gavin, V., Coleman, B., Levine, S., Nielsen, A., Carroll, L., Ursprung, S., Wood, B., Reeves, H., Keppard, B., Sportiche, N., Partirdge, J., Figueora, J., Frakt, A., Alfonzo, M., Abreu, D., Abreu, T., Ambroise, T., Andrade, E., Barrientos, E., Baty, A., Baty, C., Benner, K., Bennett, C., Blanchette, A., Bongiovanni, R., Cardile, O., Corchado, C., Dixon, C., Dodson, C., Dominguez, J., Durena, M., Fiestas, Y., Genty, J., Graffam, N., Gonzalez, A., Grigsby, E., Hayden, P., Alvado, S. H., Hernandez, Z., Hodes, I., Johnson, J., Keefe, K., Latimer, K., Logg, C., Martinez, N., Mboup, K., McPhorson, D., Meacham, S., Mohammed, D., Moss, E., O'Brien, K., Owens, L., Partridge, J., Johnson, L. P., Power, M. B., Rebelo, T., Remy, R., Roderigues, G., Sabtow, Q., Sanchez, C., Seeder, A., Sepulveda, R., West, E., Winters, L. and Youmans, T. (2018) 'Community change

and resident needs: Designing a Participatory Action Research study in Metropolitan Boston', *Health and Place*, 52(July), pp. 221–230. doi: 10.1016/j.healthplace.2018.05.014.

Audrey, S. and Batista-Ferrer, H. (2015) 'Healthy urban environments for children and young people: A systematic review of intervention studies', *Health & Place*, 36, pp. 97–117. doi: https://doi.org/10.1016/j.healthplace.2015.09.004.

Balan, P., Gerits, A. and Vanduffel, W. (2014) 'A practical application of text mining to literature on cognitive rehabilitation and enhancement through neurostimulation', *Frontiers in Systems Neuroscience*, 8, p. 182.

Blei, D., Ng, A. and Jordan, M. (2003) 'Latent Dirichlet Allocation', *Journal of Machine Learning Research*, 3, pp. 993–1022.

Cheung, N. W. T. (2014) 'Social stress, locality of social ties and mental well-being: The case of rural migrant adolescents in urban China', *Health & Place*, 27, pp. 142–154. doi: https://doi.org/10.1016/j.healthplace.2014.01.013.

Coleman, T. and Kearns, R. (2015) 'The role of bluespaces in experiencing place, aging and wellbeing: Insights from Waiheke Island, New Zealand', *Health & Place*, 35, pp. 206–217. doi: https://doi.org/10.1016/j.healthplace.2014.09.016.

Congdon, P. (1997) 'Bayesian models for spatial incidence: a case study of suicide using the BUGS program', *Health & Place*, 3(4), pp. 229–247. doi: https://doi.org/10.1016/S1353-8292(97)00017-8.

Congdon, P. (1999) 'Primary care needs assessment and resourcing: complementary practice and geographic perspectives', *Health & Place*, 5(1), pp. 59–82. doi: https://doi.org/10.1016/S1353-8292(98)00041-0.

Connelly, R., Playford, C., Gayle, V. and Dibben, C. (2016) 'The role of administrative data in the big data revolution in social science research', *Social Science Research*, 59, pp. 1–12.

Cusack, L., Larkin, A., Carozza, S. E. and Hystad, P. (2017) 'Associations between multiple green space measures and birth weight across two US cities', *Health & Place*, 47, pp. 36–43. doi: https://doi.org/10.1016/j.healthplace.2017.07.002.

Decker, M. J., Isquick, S., Tilley, L., Zhi, Q., Gutman, A., Luong, W. and Brindis, C. D. (2018) 'Neighborhoods matter. A systematic review of neighborhood characteristics and adolescent reproductive health outcomes', *Health & Place*, 54, pp. 178–190. doi: https://doi.org/10.1016/j.healthplace.2018.09.001.

Dooris, M. (2013) 'Expert voices for change: Bridging the silos—towards healthy and sustainable settings for the 21st century', *Health & Place*, 20, pp. 39–50. doi: https://doi.org/10.1016/j.healthplace.2012.11.009.

Elgar, F. J., Davis, C. G., Wohl, M. J., Trites, S. J., Zelenski, J. M. and Martin, M. S. (2011) 'Social capital, health and life satisfaction in 50 countries', *Health & Place*, 17(5), pp. 1044–1053. doi: https://doi.org/10.1016/j.healthplace.2011.06.010.

Ferry, G. (2015) 'Medical periodicals: mining the past', *The Lancet*. Elsevier, 385(9987), pp. 2569–2570. doi: 10.1016/S0140-6736(15)61151-5.

Fotso, J.-C. and Kuate-Defo, B. (2005) 'Socioeconomic inequalities in early childhood malnutrition and morbidity: modification of the household-level effects by the community

SES', *Health & Place*, 11(3), pp. 205–225. doi: https://doi.org/10.1016/j.healthplace.2004.06.004.

GBD (2017) 'Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016', *The Lancet*, 390, pp. 1151–1210.

GBD (2018) 'Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Stu', *The Lancet*, 392, pp. 1923–1994.

Jongsthapongpanth, A. and Bagchi-Sen, S. (2010) 'Spatial and sex differences in AIDS mortality in Chiang Rai, Thailand', *Health & Place*, 16(6), pp. 1084–1093. doi: https://doi.org/10.1016/j.healthplace.2010.06.014.

Kearns, R. (1993) 'Place and Health: Towards a Reformed Medical Geography', *The Professional Geographer*, 45(2), pp. 139–147.

Kearns, R. and Moon, G. (2002) 'From medical to health geography: novelty, place and theory after a decade of change', *Progress in Human Geography*. SAGE Publications Ltd, 26(5), pp. 605–625. doi: 10.1191/0309132502ph389oa.

Kukafka, R., Bales, M. E., Burkhardt, A. and Friedman, C. (2006) 'Human and automated coding of rehabilitation discharge summaries according to the International Classification of Functioning, Disability, and Health', *Journal of the American Medical Informatics Association : JAMIA*. American Medical Informatics Association, 13(5), pp. 508–515. doi: 10.1197/jamia.M2107.

Kümpers, S., Mur, I., Hardy, B., van Raak, A. and Maarse, H. (2006) 'Integrating dementia care in England and The Netherlands: Four comparative local case studies', *Health & Place*, 12(4), pp. 404–420. doi: https://doi.org/10.1016/j.healthplace.2005.04.001.

Lewis, N. D. and Rapaport, M. (1995) 'In a sea of change: health transitions in the Pacific', *Health & Place*, 1(4), pp. 211–226. doi: https://doi.org/10.1016/1353-8292(95)00030-5.

Liu, B. (2012) *Sentiment Analysis and Opinion Mining*. California: Morgan & Claypool Publishers.

Lytle, L. A. and Sokol, R. L. (2017) 'Measures of the food environment: A systematic review of the field, 2007–2015', *Health & Place*, 44, pp. 18–34. doi: https://doi.org/10.1016/j.healthplace.2016.12.007.

Madge, C. (1998) 'Therapeutic landscapes of the Jola, The Gambia, West Africa', *Health & Place*, 4(4), pp. 293–311. doi: https://doi.org/10.1016/S1353-8292(98)00033-1.

Metsker, O., Bolgova, E., Yakovlev, A., Funkner, A. and Kovalchuk, S. (2017) 'Pattern-based Mining in Electronic Health Records for Complex Clinical Process Analysis', *Procedia Computer Science*, 119, pp. 197–206. doi: https://doi.org/10.1016/j.procs.2017.11.177.

Moon, G. (1995) '(Re)placing research on health and health care', *Health & Place*, 1(1), pp. 1–4.

Moon, G. (2008) 'Enhancing immediacy', *Health & Place*, 14(1), p. 1. doi: https://doi.org/10.1016/j.healthplace.2007.10.005.

Moon, G. and Pearce, J. (2020) 'Twenty-five years of Health & Place: citation classics, internationalism and interdisciplinarity', *Health & Place*.

Ning, W., Chan, S., Beam, A., Yu, M., Geva, A., Liao, K., Mullen, M., Mandl, K. D., Kohane, I., Cai, T. and Yu, S. (2019) 'Feature extraction for phenotyping from semantic and knowledge resources', *Journal of Biomedical Informatics*, 91, p. 103122. doi: https://doi.org/10.1016/j.jbi.2019.103122.

Oakes, J. M., Andrade, K. E., Biyoow, I. M. and Cowan, L. T. (2015) 'Twenty Years of Neighborhood Effect Research: An Assessment', *Current Epidemiology Reports*, 2, pp. 80–87. doi: 10.1007/s40471-015-0035-7.

Philo, C. (1997) 'Across the water: Reviewing geographical studies of asylums and other mental health facilities', *Health & Place*, 3(2), pp. 73–89. doi: https://doi.org/10.1016/S1353-8292(97)00002-6.

Porter, C., Atkinson, P. and Gregory, I. (2015) 'Geographical Text Analysis: A new approach to understanding nineteenth-century mortality', *Health & Place*, 36, pp. 25–34. doi: https://doi.org/10.1016/j.healthplace.2015.08.010.

Richardson, D., Volkow, N., Kwan, M.-P., Kaplan, R. and Goodchield, M. (2013) 'Spatial Turn in Health Research', *Science*, 339(6126), pp. 1390–1392.

Silge, J. and Robinson, D. (2017) *Text Mining with R: A Tidy Approach*. California: O'Reilly.

Spasić, I., Livsey, J., Keane, J. A. and Nenadić, G. (2014) 'Text mining of cancer-related information: Review of current status and future directions', *International Journal of Medical Informatics*, 83(9), pp. 605–623. doi: https://doi.org/10.1016/j.ijmedinf.2014.06.009.

Timmins, K., Green, M. A., Radley, D., Morris, M. A. and Pearce, J. (2018) 'How has big data contributed to obesity research? A review of the literature', *International Journal of Obesity*.

Tunstall, H., Cabieses, B. and Shaw, R. (2012) 'The characteristics of mobile families with young children in England and the impact of their moves on neighbourhood inequalities in maternal and child health', *Health & Place*, 18(3), pp. 657–670. doi: https://doi.org/10.1016/j.healthplace.2011.11.009.

Westergaard, D., Stærfeldt, H.-H., Tønsberg, C., Jensen, L. and Brunak, S. (2018) 'A comprehensive and quantitative comparison of text-mining in 15 million full-text articles versus their corresponding abstracts', *PLoS Computational Biology*, 14(2), p. e1005962.

Wilton, R. D. (1996) 'Diminished worlds? The geography of everyday life with HIV/AIDS', *Health & Place*, 2(2), pp. 69–83. doi: https://doi.org/10.1016/1353-8292(95)00040-2.